Computational Design of Biomimetic Phosphate Scavengers

Gruber, Mathias Felix; Wood, Elizabeth Baker; Truelsen, Sigurd Friis; Østergaard, Thomas; Hélix-Nielsen, Claus

Published in:
Environmental Science and Technology

Link to article, DOI:
10.1021/es506214c

Publication date:
2015

Document Version
Peer reviewed version

Computational Design of Biomimetic Phosphate Scavengers

Mathias F. Gruber[†], *Elizabeth Wood*[†], *Sigurd Truelsen*[†], *Thomas Østergaard*[†], *and Claus Hélix-Nielsen*[†‡*]

[†] The Biomimetic Membrane Group, Department of Environmental Engineering, Technical University of Denmark, DK 2800 Kgs. Lyngby Denmark

[‡] University of Maribor, Faculty of Chemistry and Chemical Engineering, Smetanova ulica 17, SI-2000 Maribor, Slovenia

KEYWORDS: phosphorus recovery, intrinsically disordered proteins, molecular dynamics, enhanced sampling

***) Corresponding Author**

Claus Helix-Nielsen

Department of Environmental Engineering

Building 115, office 140,

Technical University of Denmark

DK2800 Kgs. Lyngby, Denmark

EMB: **clhe@env.dtu.dk**

TEL: +45 45 25 22 28

FAX: +45 45 93 28 50

1 Abstract

Phosphorous has long been the target of much research, but in recent years the focus has shifted from being limited only to reducing its detrimental environmental impact, to also looking at how it is linked to the global food security. Therefore, the interest in finding novel techniques for phosphorous recovery, as well as improving existing techniques, has increased. In this study we apply a hybrid simulation approach of molecular dynamics and quantum mechanics to investigate the binding modes of phosphate anions by a small intrinsically disordered peptide. Our results confirm that the conformational ensemble of the peptide is significantly changed, or stabilized, by the binding of phosphate anions and that binding does not take place purely as a result of a stable P-loop binding nest, but rather that multiple binding modes may be involved. Such small synthetic peptides capable of binding phosphate could be the starting point of new novel technological approaches towards phosphorus recovery, and they represent an excellent model system for investigating the nature and dynamics of functional de novo designed intrinsically disordered proteins.

2 Introduction

Phosphorous (P) is an essential element in terms of sustaining the world's current and future food supply, for which there is no substitute.[1–3] Given that the current P supply is based on the gradual depletion of limited fossil reserves, an increasing demand for P necessitates a change towards more sustainable practices where P is recovered from the large waste streams. The lifetime of remaining high quality phosphate rocks is still being debated, estimates varying from a few decades to a few hundred years.[3,4] There is however a general consensus that P is becoming more and more difficult to access, costs are increasing, more waste is being produced, and the

global demand is expected to increase.[4,5] Meanwhile, only a fraction of the mined P makes it into the intended plants and animals which humans consume, while most is lost along the way causing serious environmental problems e.g. by eutrophication of lakes, reservoirs, estuaries, and parts of the ocean.[2,4,6,7]

The topic of P has been a point of interest for waste-water treatment engineers for decades.[8] The main attention has however so far been focused almost exclusively on reducing eutrophication, so while many of the now common techniques for P treatment, e.g. chemical precipitation[8] and enhanced biological phosphorus removal[9] (EBPR), are highly efficient for the job they were designed for, they are not necessarily effective in terms of recovering P from its large waste flows, which have different characteristics from the commonly treated domestic wastewater flows and are not always easily intercepted (e.g. erosion and runoff[2]). One of the current main technologies, optimized for P removal but also applicable to recovery to some extent, is EBPR, where polyphosphate accumulating organisms are used to capture and store high amounts of P in their heterotrophic biomass. These organic biosolids may subsequently need to be treated prior to reuse of the P, or they may be used directly in agricultural settings, e.g. as a slow-release fertilizer. EBPR however faces several limitations, perhaps the most serious being that energy recovery (methane production) must be carried out in a strictly anaerobic system, which is not easily combined with the aerobic and anaerobic conditions needed in EBPR-based waster water treatment.[2,10]

Biological techniques such as EBPR are encouraged by the fact that certain microorganisms flourish in P-limited environments by having developed efficient enzymes and proteins that bind with high specificity and reversibility to phosphorus compounds.[11,12] This makes biomimetic approaches for P recovery generally interesting, and we recently did a

statistical analysis of how proteins in nature bind different phosphorus compounds in order to reveal common binding site characteristics.[13] One of the most common ways in which proteins bind phosphates non-covalently, and in particular the β-phosphate of ATP and GTP, is through the P-loop, which is characterized by the consensus sequence Gly-Xxx-Xxx-Xxx-Xxx-Gly-Lys-(Ser,Thr).[14–18] Inspection of the three-dimensional structure of P-loops reveals that it is remarkably well conserved throughout nature, and the main conformation of the sequence is generally found to form a feature resembling an anion-binding nest.[19,20] A classic "nest" feature in biochemistry is defined to consist of three amino acids, and can be either in a LR or RL configuration, depending on the main chain dihedral angles of the two first residues.[21] The anion-binding nest seen for the P-loops is formally a series of overlapping nests, typically in an LRLR conformation for the Xxx-Xxx-Gly-Lys part of the consensus sequence.[22] Although the P-loop nest structure is expected to be essential for binding of the phosphate anion, the anion is also expected to stabilize the nest conformation, which has been demonstrated for some P-loop proteins by X-ray crystallography; i.e. the P-loop without the anion is supposedly fairly flexible and without any stable structure.[23,24]

Recently, a hexapeptide with the sequence Ser-Gly-Ala-Gly-Lys-Thr (SGAGKT) was synthesized Bianchi *et al*. based on the P-loop consensus sequence.[25] The two glycines in this peptide were introduced to promote a natural LRLR conformation, and the peptide in its lysine-protonated zwitter-ionic state was found be a capable binder of phosphate anions PO_4^{3-} and HPO_4^{2-}, whereas $H_2PO_4^{1-}$ and H_3PO_4 were found not to be bound. From their studies, it is expected that the synthetic peptide form a classical LRLR nest structure when binding of the anion. Due to its small size, it is reasonable to assume that when unbound in solution this peptide does not posses a well-defined secondary structure and instead exists as a dynamic ensemble of conformations which may posses transient residual secondary structure, or consist of multiple structures that rapidly

exchange.[26] As such it is not unreasonable to assume that the peptide might potentially have a nature similar to intrinsically disordered proteins (IDPs), IDPs having received increasing interest in the scientific community in recent years,[27] given that they compose approximately 20% of the proteome.[28] Conventionally, when a given ligand binds to an IDP, the favorable free energy from binding is offset by the loss in conformational entropy, resulting in complexes that can be highly specific with low overall binding energies.[29] For the SGAGKT peptide, however, binding of the phosphate anions has experimentally been found to be attributed to favorable entropic contributions.[25]

In order to further investigate the nature of how the P-loop binds and interacts with phosphate, in this study we apply molecular dynamics (MD), accelerated molecular dynamics (aMD) and semi-empirical quantum mechanics (QM) in a hybrid approach (QM/MM) to simulate how the phosphate anion is bound by the peptide SGAGKT, and demonstrate how these results correlate with experimental data. MD is a simulation technique which can be used for simulating the physical movements of atoms in molecules in the context of classical dynamics, using empirically determined forcefields.[30] For more details on the approach see Appendix A. Our simulations confirm the suspected intrinsically disordered nature of the peptide in the absence of an anion, and demonstrate how the addition of an anion stabilizes the conformational space. They also suggest, however, that the phosphate anions are not completely stably bound in a singular P-loop nest structure, as hypothesized in literature, but rather that multiple binding conformations exist; this translates to a very reversible binding mechanism, and is therefore of considerable interest for applications involving phosphorus recovery. The use of a phosphate-binding peptide in a biomimetic system represents a model where there is no need for growing heterotrophic bacteria. Despite such technology still being at its infant state, it may thus potentially be superior

to current technologies such as EBPR, and it is therefore an appealing area of research in the future of resource recovery.

3 Materials and Methods

3.1 Model System

All MD simulations were run using the molecular modeling packages Amber14 and AmberTools14.[30] To set up a given peptide simulation, first the fully extended peptide was created using the module LEaP in AmberTools14. For peptide-anion systems, the anion was added randomly to the system on a sphere 100Å away from the peptide, hereafter it was translated to 3.5Å proximity of the closest atom in the peptide, thereby ensuring random starting positions. All systems were built with LEaP using the ff14SB force field parameters[31] and were explicitly solvated using the TIP3P model for water[32] in rectangular boxes with 10.0Å to each edge from the system of interest. This resulted in each simulation having between 1450 and 1650 water molecules. To minimize system complexity no counter ions were included in the simulations.

The systems were minimized for 10,000 steps: 5,000 steps using a steepest descent algorithm and 5,000 steps using a conjugated gradient algorithm. Heating of the system to 300K over 0.8ns was then performed with a weak restraint of 1 kcal/mol on the peptide-anion, and afterwards the systems were equilibrated for 0.2 ns without any restraints at 300K. Heating was performed under constant volume and temperature (NVT ensemble). Equilibration and production runs were performed at 300K using the Langevin thermostat with a collision frequency of 1 ps^{-1}, and the pressure was set to 1 bar using the Berendsen barostat (NPT ensemble). Hydrogens were restrained using SHAKE[33], non-bonded interactions were cut off at

10.0 Å, and full electrostatics for the periodic system were calculated using the Particle Mesh Ewald approach.[34]

In all peptide-anion simulations, the phosphate anions were marked as quantum regions and treated using the semi-empirical PM6 hamiltonian with dispersion correction as implemented in Amber14.[30,35] SHAKE restriction was used on all hydrogens in the QM region and all simulations were performed with a 2 fs timestep. The *sander* module of Amber14 was used for QM/MM simulations and *pmemd.cuda* was used for all peptide-only simulations.

3.2 aMD simulations

For all aMD simulations the average potential and dihedral energies were first obtained from 2 ns cMD runs on the same system. The hexapeptide SGAGKT has 74 atoms, and with the approximation that each residue has an energy contribution of 3.5 kcal/mol, the values for the parameters required for the aMD runs (see Appendix A) were calculated using empirical estimates as specified in Eqs. 1-4, where αD and αP are added to the average values obtained from the cMD simulations in order to account for the degrees of freedom in the peptide.[36]

Equation 1

$$\alpha D = \frac{1}{5} \cdot 6 \cdot 3.5 \frac{\text{kcal}}{\text{mol}} = 4.2 \frac{\text{kcal}}{\text{mol}}$$

Equation 2

$$Ed = Ed_{\text{avg}} + \alpha D$$

Equation 3

$$\alpha P = \frac{1}{5}\frac{\text{kcal}}{\text{mol}\cdot\text{atoms}} \cdot 28\ \text{atoms} = 5.6\frac{\text{kcal}}{\text{mol}}$$

Equation 4

$$Ep = Ep_{\text{avg}} + \alpha P$$

To obtain the canonical ensemble the trajectories were reweighted through cumulant expansion to the second order, using the python toolkit provided McCammon group.[37]

3.3 Cluster Analysis & Principal Component Analysis

Clustering of trajectories and principal component analysis (PCA) were performed with CPPTRAJ, which is the native post-processing utility of Amber14.[38] Clustering was done both with an agglomerative hierarchical algorithm with ε set to 3.0 and using the DBSCAN[39] algorithm with ε set to 1.2 and the minimum points set to 20. The distance metric used for both algorithms was the root mean square deviation (RMSd) of all atoms in the peptide except for hydrogens. To reduce processing time, a "sieve" value was chosen such that for a given trajectory, only 20,000 evenly spaced frames from the trajectory were used, and afterwards sieved frames were added back into the clusters if applicable given the value of ε.

A powerful way to improve the information gained from clustering algorithms, is to combine it with principal component analysis (PCA)[40], where the overall motions of the trajectories are represented in a lower dimensional space by mapping the trajectories to a set of eigenvectors calculated from the covariance matrix of the atoms.[41,42] These eigenvectors are referred to as "modes" throughout this work, such that each mode corresponds to a certain type of motion of the

peptide. Prior to all PCA the trajectories were RMS fitted to the overall average structure first, and the coordinate covariance matrix was then calculated for all heavy atoms in the peptide. The trajectories were projected along the calculated modes in order to obtain a time series of PCA projection values for each mode.

The Kullback-Leibler divergence (KLD) has previously been shown to be a good indicator of convergence between two independent simulations.[43–46] Briefly, the KLD can be defined as

Equation 5

$$KLD(t) = \sum_i P(t,i) \cdot \ln\left(\frac{P(t,i)}{Q(t,i)}\right)$$

where $P(t,i)$ and $Q(t,i)$ are the probability distributions of two independent simulations with t being the time and i being the bin index. Practically, the KLD is a measure of information difference between the two probability distributions, and as such when it converges towards zero the distributions can be said to have converged. To calculate the KLD between two simulations, PCA was first performed on the combined trajectory of the two simulations, and then the normalized probability distributions along the first (largest) eigenvector for each simulation were used in Eq. 5. For all histograms a total of 400 bins were used and a Gaussian kernel density estimator, a fundamental method for estimating probability density / data smoothing, was used to reduce the amount of bins with no population.[43]

Structure alignment was done using the Kabsch algorithm, which is a method for calculating the optional rotation matrix that minimize the RMSd between two sets of points.[47] As such it can be used to align peptide structures and superposition them on top of each other.

4 Results & Discussion

4.1 Sampling the peptide conformational ensemble

Peptide-only cMD simulations were performed using the GPU-accelerated pmemd.cuda module of Amber14, which enabled simulation times for each simulation of $100 - 200$ ns/day with a single NVidia Tesla M2050 node. To test whether the conformational ensemble for these simulations had converged, KLD between the major modes of a PCA decomposition of a combined trajectory for two independent simulations was calculated.

In Fig 1a, the KLD of the first three PCA modes are shown for the two peptide-only simulations, confirming that after around 1 μs the value is less than 0.02, which is used as the threshold value for declaring convergence.[43,45] Correspondingly, in Fig. 1b the RMSd frequencies for these two simulations are shown after the 1 μs of simulation; i.e. it is a normalized frequency plot of RMSd values, where the RMSd values have been calculated between each frame in the trajectories and a common reference structure. Based on Fig. 1 it can be concluded that the conformational ensemble of the peptide-only simulations are well-converged after 1 μs of cMD simulation, however it is noted how the KLD of mode 1 and 2 increases drastically at around 400 ns which can be attributed to the two simulations exploring different parts of the conformational space at that point. This highlights the dangers of interpreting too short simulations with KLD. It was visually confirmed from the trajectories that during the simulation, the peptide undergoes such large changes in conformation that additional cMD runs are deemed unlikely to change the conformational space investigated or the obtained ensemble.

4.2 Restriction of the conformational space by HPO_4^{2-}

Introducing the phosphate anion into the system means it is no longer possible to use the GPU implementation of pmemd in Amber14, since QM is not currently supported in the pmemd modules. As a result, simulation times were limited to approximately $5-7$ ns/day for each simulation when using 16 cpu cores (Intel Xeon E5-2680, 2.80 GHz) with the sander module of Amber14. Four independent simulations were carried out with the peptide$-HPO_4^{2-}$ system for 100 ns. In Fig 2, 2d RMSd plots for a simulation of the peptide$-HPO_4^{2-}$ system, and a simulation of a peptide-only system are shown. Despite the fact that neither of these systems are converged after 100 ns, the seen behavior was consistently observed for all four anion simulations: the conformations sampled by the peptide throughout the simulation in the peptide$-HPO_4^{2-}$ systems are retained for longer periods of time and appear more "stable" on the 100 ns timescale, whereas the conformations sampled in the peptide-only simulations change more rapidly.

4.3 Enhanced sampling with aMD

The issue of poor conformational sampling of multiple-molecule systems in MD is a long-standing problem, and several attempts have been made to use different enhanced sampling techniques to describe atomistic systems with varying levels of success. Here we use aMD to see if we can improve the sampling time of the canonical ensemble for both peptide-only and peptide$-HPO_4^{2-}$ simulations. In the case of peptide-only simulations, it was found that the KLD for two aMD simulations go below 0.02 already after 50-100 ns, and is further reduced to $\sim 10^{-3}$ after the full 1 μs run (see supplementary information, Fig. S1). Considering that for cMD simulations convergence was declared only after 700 ns, aMD clearly represents a powerful way of increasing the sampling time for the conformational ensemble of the peptide-only simulations. For it to be truly useful, reweighting must be performed to obtain the original ensemble – using a cumulant

expansion algorithm to the second order, this reweighting was found to successfully reproduce the canonical ensembles from 50 ns aMD peptide-only simulations (see supplementary information, Fig. S2). The challenge, then, is whether or not aMD can similarly be used to simulate converged ensembles for the peptide-anion systems investigated in this work as well.

In Fig. 3 the KLD's between four independent 100 ns cMD and aMD peptide$-HPO_4^{2-}$ simulations are shown. The KLD for the cMD peptide $- HPO_4^{2-}$ simulations are not seen to decrease notably, meaning that the conformational ensembles have not converged during the 100 ns of simulation time. Compared to the cMD simulations, it is clear from Fig. 3 that aMD increases sampling speed, but also that despite the relatively large boost from aMD (avg. 4.5 kcal/mol), it does not reach full convergence ($KLD < 0.02$) after 100 ns. Given that it takes the conformational ensembles of peptide-only simulations on the order of 1 μs to converge using cMD and 50-100 ns using aMD, it is no surprise that a peptide-anion system does not converge during 100 ns. The introduction of an anion into the system significantly complicates the system and thereby the required time for the ensemble to converge; it is noted that the ligand will likely also spend a certain amount of time in an unbound state, such that the convergence time should be at least that of the peptide-only system. One can only speculate about the time-scale required for the ensemble of the peptide$-HPO_4^{2-}$ system to converge fully, but it is likely to be on the order of several micro-seconds if not even in the milli-second range for cMD and micro-seconds for aMD.

Recalling that a KLD plot for two simulations is created from the overlap between the PCA histograms, KLD values of 0.1 in the case of the aMD peptide$-HPO_4^{2-}$ system clearly indicate that the four simulations presented in Fig. 3 sample similar motions and conformations and there is a significant overlap. As such, these close-to-converged ensembles can still be analyzed in terms of

259 which conformations are present, with the precaution that the true frequency of each
260 conformation in the ensemble is unknown.

261 ## 4.4 Cluster & PCA analyses

262 MD trajectories can contain on the order of millions of frames, so in order to obtain a
263 qualitative picture of a given system's properties, reduced representations of the trajectories must
264 be constructed first. The two most commonly used techniques for creating such reduced
265 representations are principal component analysis (PCA) and clustering algorithms.[40]

266 Clustering algorithms can be separated into two basic types; hierarchical and partitioning
267 algorithms.[48] In hierarchical algorithms the decomposition can be viewed in the form of a
268 dendrogram, i.e. a tree where the dataset *D* is split into smaller subsets such that each node of the
269 tree represents a cluster. This can be done either from the leaves up (*agglomerative approach*) or
270 from the root down (*divisive approach*). Regardless of the approach, the hierarchical algorithms
271 require the input of a termination condition, e.g. a critical distance between each cluster – this is
272 the main problem with those kind of algorithms, since the clusters are sensitive to small changes
273 in the termination condition as well as noise in the dataset. With partitioning algorithms the
274 dataset *D* is initially split into a set of *k* clusters and then an iterative strategy is used to optimize
275 some objective function. These algorithms require *k* as an input parameter, which limits their use
276 since enough knowledge about the domain may not be known beforehand.

277 An alternative clustering approach to the two basic types are density based clustering
278 algorithms, where clustering is performed based on definitions of densities and connectivities in
279 the dataset.[39,49,50] One of the most common of these is the DBSCAN algorithm[39], which creates
280 clusters based on a simple notion of density-connectivity between points in the dataset. The
281 DBSCAN algorithm requires the input parameter for minimum points in a given cluster and ε

which characterizes the ε-neighbourhood of a given point, i.e. the connectivity of points in the cluster. DBSCAN filters out points in the dataset that do not belong in a given cluster as "noise", which would otherwise be added into the closest cluster in other algorithms (e.g. the hierarchical). It furthermore supports an effective heuristic denoted the *sorted k-dist graph,* i.e. a sorted list of the Kth farthest distance for each point in the dataset, which helps the user in determining the two input parameters.[39] Clustering is inherently a highly complicated task, and the DBSCAN algorithm suffers from several disadvantages, e.g. it expects a density drop to detect the borders of the clusters, which means it might not be able to detect some of the more intrinsic clusters present in natural dataset. The DBSCAN algorithm has been revisited in the form of several extensions and modifications since its first description[51–53], however the algorithm in its original form has stood the test-of-time and is generally considered a powerful clustering algorithm.

To determine binding modes for the SGAGKT peptide to phosphate anions, we used both a hierarchical agglomerative algorithm and the DBSCAN algorithm. The cutoff used for the hierarchical algorithm was set to 3Å, which was empirically determined to result in approximately 10-20 clusters for each simulation. For DBSCAN the minimum points and ε parameters were set to approximately 20 and 1.2 respectively, based on k-dist plots (See supplementary information, Fig. S3).

In Fig. 4, clustering results are presented for peptide-only (1 μs cMD) along the PCA projections for the two major modes (comprising $\sim$ 60% of the total peptide motion, see supplementary information, Fig. S4-S6). It is clearly observed how the two different clustering algorithms are different in terms of how they cluster the trajectories. The DBSCAN algorithm sorts out a large part of the trajectory as noise (78 ± 1% for peptide-only simulations), whereas the

hierarchical algorithm clusters all points together based on their closeness to each other. The peptide bound to HPO_4^{2-} does show transiently stable conformations as evident from the DBSCAN analysis (see supporting information, Fig. S7), however, the majority of the time is spent in conformations of a more disordered nature ($57 \pm 10\%$ is filtered off as noise), at least within the DBSCAN terminology and the algorithm input parameters. The same is evident from the free energy profiles, where local minima are observed for both the peptide-only and peptide-HPO_4^{2-} simulations; the barriers around these minima are however fairly broad and smooth, which reflects the disordered nature of the peptide.

In Fig. 5 aligned superpositions of the different clusters are presented for four 100 ns peptide-HPO_4^{2-} simulations and two 1 μs peptide-only simulations, which is done to visualize the difference between the semi-stable DBSCAN conformations and the hierarchical clusters of the peptide. In Fig. 5a-b all the DBSCAN clusters, containing both peptide and anion, are shown – it is evident that these have a tendency towards the expected P-loop structure, where the backbone and the lysine side-chain folds around the anion in a nest structure. In the case of peptide$-HPO_4^{2-}$ simulations, $57 \pm 10\%$ of the ensemble was filtered off as noise by the DBSCAN algorithm: compared to the $78 \pm 1\%$ for peptide-only simulations, this again shows that the anion stabilizes the disordered ensemble. The hierarchical clusters shown in Fig. 5c-d on the other hand display much more variation in the peptide-anion interaction: in a significant amount of the clusters the anion is found to be interacting with the more or less extended peptide by only 1-3 hydrogen bonds without any nest-like structure. For the peptide-only simulations the DBSCAN clusters in a similar fashion reveal a series of semi-stable nest-like states (Fig. 5e), where the backbone is folded up in a nest with the lysine chain is in a more or less indeterminate orientation. The hierarchical clusters on the other hand show the disordered nature of the peptide – and it is in this disordered state that the peptide spends most of its time (Fig. 5f).

In the original paper where Bianchi *et al.* synthesized and investigated SGAGKT experimentally, they found that the peptide in its lysine-protonated zwitterionic state bound to HPO_4^{2-} with $\Delta G = -4 \pm 0.1$ kcal/mol, whereas $H_2PO_4^{1-}$ was found not to be bound. The ensembles of our peptide$-HPO_4^{2-}$ simulations are not fully converged, but can still be considered close-to-converged. It is therefore interesting to see how these ensembles compare against the experimental binding energy; this can be done using a method known as molecular mechanics Poisson-Boltzmann surface area (MM-PBSA), which is a post-processing approach to estimating free energies and binding energies of molecules in solution.[54] For the combined 400 ns trajectory of the peptide$-HPO_4^{2-}$ simulations, this approach calculates an average favorable binding free energy of 1.72 ± 0.28 kcal/mol. For 200 ns peptide-$H_2PO_4^{1-}$ simulations, an average favorable binding free energy of 1.45 ± 0.24 kcal/mol was found, indicating that also this anion is bound by the peptide, albeit more weakly, which can likely be attributed to the less electronegative nature of $H_2PO_4^{1-}$. Looking at the binding energy distributions obtained from MM-PBSA (see supplementary information, Fig. S8), it is observed that these are markedly different for the two anions, with HPO_4^{2-} having a broader distribution that is shifted towards more favorable binding energies compared to $H_2PO_4^{1-}$ which is more centered around $\Delta G = 0$, and it is clear from the energy distributions that there is a difference between how $H_2PO_4^{1-}$ and HPO_4^{2-} bind to the peptide. The discrepancy with experimental results can be attributed to the close-to-converged nature of the simulations, the inadequacy of MM-PBSA in describing binding energies in such highly dynamic systems, as well as the presence of 1M NMe_4Cl in the experimental setup; such additional ions are likely to influence the system in a way that is not accounted for in the present theoretical model. The DBSCAN algorithm furthermore reveals that that $84 \pm 3\%$ of the trajectory for the peptide$-H_2PO_4^{1-}$ simulations is filtered off as noise, indicating significant less stable

structures for this system compared to peptide$-HPO_4^{2-}$. Looking at the hierarchical clusters for the peptide$-H_2PO_4^{1-}$ simulations, many of them represent states in which the anion is not bound (see supplementary information, Fig. S9). Altogether, simulations qualitatively show that there is a clear difference between how $H_2PO_4^{1-}$ and HPO_4^{2-} are bound by the peptide, which is consistent with the experimental observations.

High-resolution simulations such as the ones presented in this study are very difficult to interpret, since in addition to their inherent approximations in the form of choice of force field parameterization etc, their complexity necessitates the use of enhanced sampling techniques in order to reach convergence of the molecular ensembles, which in turn bias the obtained results. Despite these limitations, such simulations do provide insight into the dynamics at the molecular scale that can otherwise be difficult to obtain experimentally. Using accelerated molecular dynamics, which implicitly conserves the overall energy landscape of the system, we are able to obtain fully-converged peptide-only ensembles, and close-to-converged peptide-anion ensembles, which can then be reweighted to the canonical ensembles. The expected intrinsically disordered conformation of the peptide is confirmed from the simulations alongside with a transiently semi-stable nest structure, and it is shown that this conformational ensemble of structures is stabilized upon HPO_4^{2-} binding, which is in accordance with expectations and theories in the literature: however, albeit there is a clear tendency for the peptide to bind the HPO_4^{2-} anion using a P-loop nest structure, slightly over half of the structures in the ensemble bind, or simply interact, with the anion in more loosely defined conformations. As such the binding process should not be considered in terms of the anion being tightly bound by the peptide in a single conformation, such as is the case for many conventional protein-ligand systems; rather it is a much more dynamic

binding effect with multiple structured states that can rapidly interchange and either bind, or release the anion.

It is important to remark that simulations were limited to the zwitterionic peptide with a protonated lysine residue, in the absence of any other ions such as Na^+ or Mg^{2+} cations. The influence of such additional species in the simulation is not immediately clear, as they may both work to facilitate binding as is often observed in nature[13], or to disturb the binding energetics of the system. Bianchi *et al.* who synthesized the SGAGKT peptide performed their experimental studies in NMe_4Cl solutions[25], indicating that in this solution the peptide is capable of binding the phosphate anion. It is however not apparent how the presence of other species may influence the enthalpy and entropy contributions to the binding affinity.

Despite the complicated binding mechanism and model approximations (protonation states, absence of counter ions, choice of force field, neglect of potential cooperative binding etc.), the theoretical model used here found that the peptide is capable of distinguishing between HPO_4^{2-} and $H_2PO_4^{1-}$, which is consistent with experimental results and shows that despite the presence of only 6 amino acid residues, the peptide has very specific binding properties. The nature of the SGAGKT peptide in terms of binding phosphate anions, i.e. its reversible binding effect and specificity, makes it very interesting for development of new technologies for P recovery, where a too strong binding may be counter-productive when it comes to the recovery process. In such applications, a dynamic/reversible binding mechanism such as the one observed for the peptide may be advantageous in terms of a subsequent recovery step.

Appendix A: Sampling Considerations

A key drawback of the classical MD approach is the assumption that the electrostatic properties of molecules can be represented using point charges at the nuclear sites.[55] In this

respect, QM provides a more rigorous treatment of the quantum chemical nature of the system at the price of a higher computational cost. For the system investigated in this report, we use QM specifically to describe the negatively charged phosphate anion, where the electronic structure is expected to be highly polarized, something which is not accounted for in the classical MD approach. For more thorough information about the benefits and limitations of the QM/MM approach in Amber14, the reader is referred reviews on the subject.[56]

One of the key challenges in MD is to obtain "adequate" sampling of the conformational space of the system, such that all-important conformational states are sampled close to their Boltzmann-weighted ones. From such well-converged ensembles, one can calculate various thermodynamic properties, and thereby validate the simulation against experimental data. Even with recent advances in computing power, which have made microsecond and even millisecond time scale available to certain researchers,[57–59] it can however still be difficult to obtain well-converged ensembles of biological molecules using MD.[45,60] The issue at hand is that the systems of interest in chemistry, physics and biology are characterized by the presence of a number of metastable states, which are separated by large barriers in the energy landscape, meaning that the system is easily trapped in a local minimum during a MD simulation. When two or more biomolecules are present, such as is the case in binding events between a IDP and its ligand, the situation is often complicated even further and generally equilibrium simulations of coupled folding and binding events at atomistic resolution are considered out-of-reach for the average researcher[61], and instead coarse-grained representations of the systems are used.[62]

In the case of atomistic simulations, various techniques to improve sampling have been explored, e.g. self-guided Langevin dynamics (SGLD)[63], accelerated MD (aMD)[64], and different variations of replica exchange MD (REMD).[65–67] The different enhanced sampling techniques all

serve to increase sampling, but each also has its own set of disadvantages; e.g. REMD in general requires running *N* non-interacting replicas of the system, which may be prohibitive, temperature REMD (T-REMD) cannot guarantee convergence since not all barriers in a system are necessarily temperature-dependent[68], reservoir REMD (R-REMD) is dependent on knowledge contained in a pre-generated reservoir of structures[60], and in SGLD, which accelerates low-frequency motion in the system, the ensemble has to be reweighted afterwards to obtain the canonical ensemble.[63] In aMD a bias potential is introduced into the conventional MD (cMD) simulation which in practice lowers the height of local energy barriers, such that the sampling can continue faster; it inherently represents an increased sampling method where only a single copy of the system is simulated, and it does not require any previous information about the energy landscape or conformation space of the system.[64] The aMD modification is defined as:

Equation 1

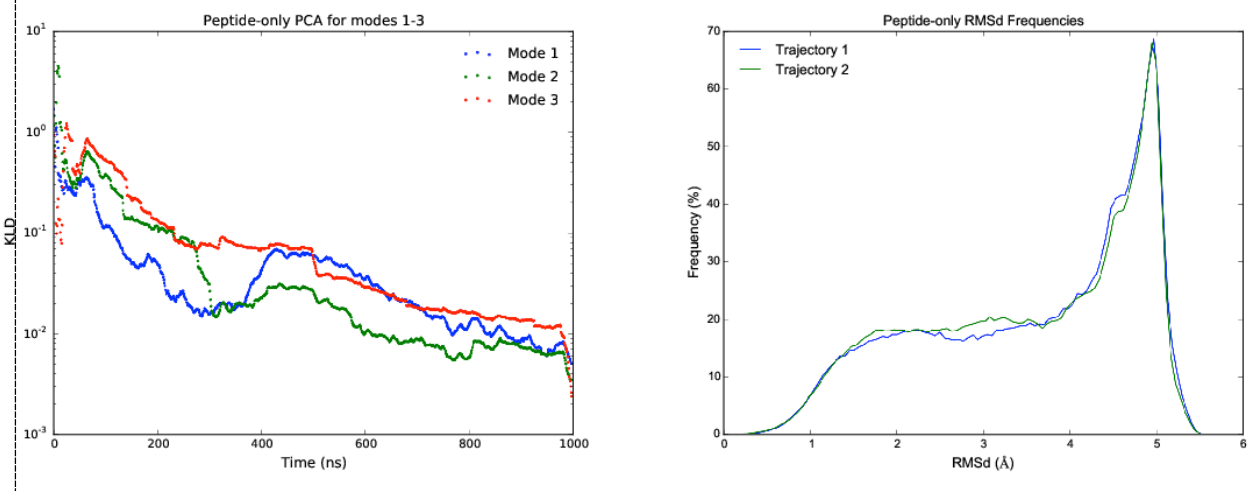
$$V(r)^* = V(r) + \Delta V(r)$$

Equation 2

$$\Delta V(r) = \frac{(Ep - V(r))^2}{(\alpha P + Ep - V(r))} + \frac{(Ed - Vd(r))^2}{(\alpha D + Ed - Vd(r))}$$

where *V(r)* is the traditional MD potential, $V(r)^*$ is the modified potential, *Vd(r)* is a torsion potential, $\Delta V(r)$ is the applied bias, *Ed* and *Ep* are the average dihedral and potential energies, and αP and αD are factors for determining the strength of the applied boost (i.e. high values reduce

441 the boost). This potential is proportionally bigger for deep regions in the energy landscape, and
442 smaller for high-energy regions, thus conserving the shape of the landscape such that minima are
443 still minima, and vice versa for barriers. This means that in theory the original canonical ensemble
444 can be recovered exactly by reweighting the distribution.[64] In practice reweighting of biased
445 ensembles can however be challenging due to statistical errors[69,70], and several algorithms for the
446 task have been proposed, see ref([37]). Previous studies have demonstrated how 500 ns aMD
447 simulations could successfully be used to recover the correct canonical ensembles when
448 compared to millisecond MD simulations and experimental data[36], truly highlighting the power of
449 aMD to obtain converged ensembles on a scale otherwise only available to a limited number of
450 researchers.
451

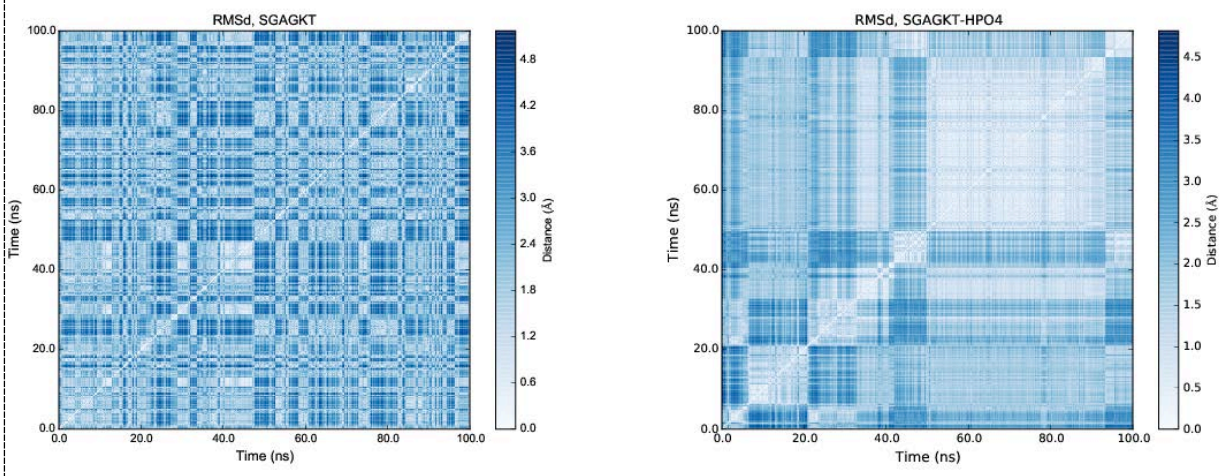


452

453 FIGURE 1: Results from peptide-only simulations in explicit water showing (a) KLD of the

454 first three PCA modes and (b) Frequency of RMSd values of frames in the trajectory

455 compared to common reference structure (the fully extended peptide).



456

457 FIGURE 2: 2d RMSd plots for 100 ns simulations with the SGAGKT peptide in the absence

458 (a) and presence (b) of the HPO_4^{2-} anion in explicit water.

459

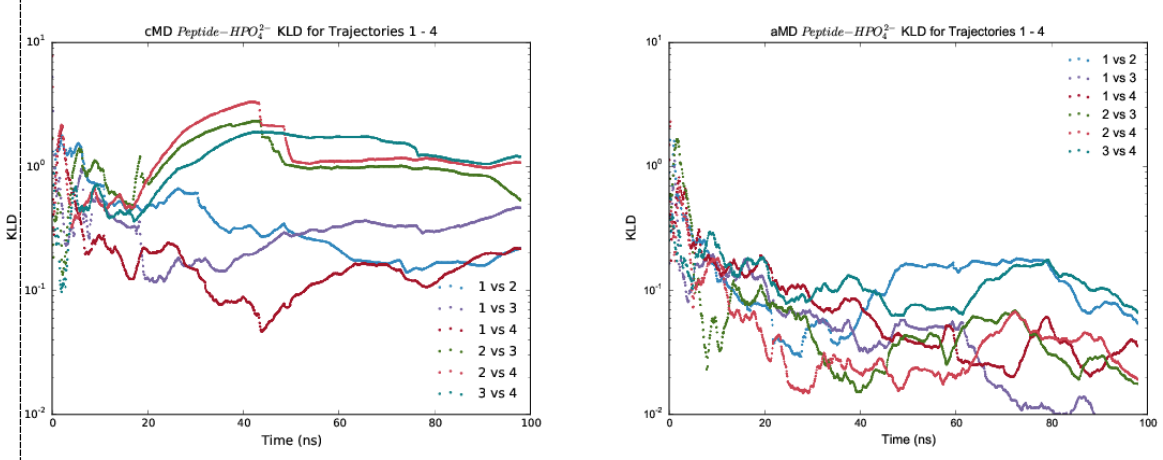


FIGURE 3: KLD of four independent cMD (a) and aMD (b) peptide-HPO_4^{2-} simuations (1-4) performed in explicit water. For each simulation the KLD is calculated against all the other simulations.

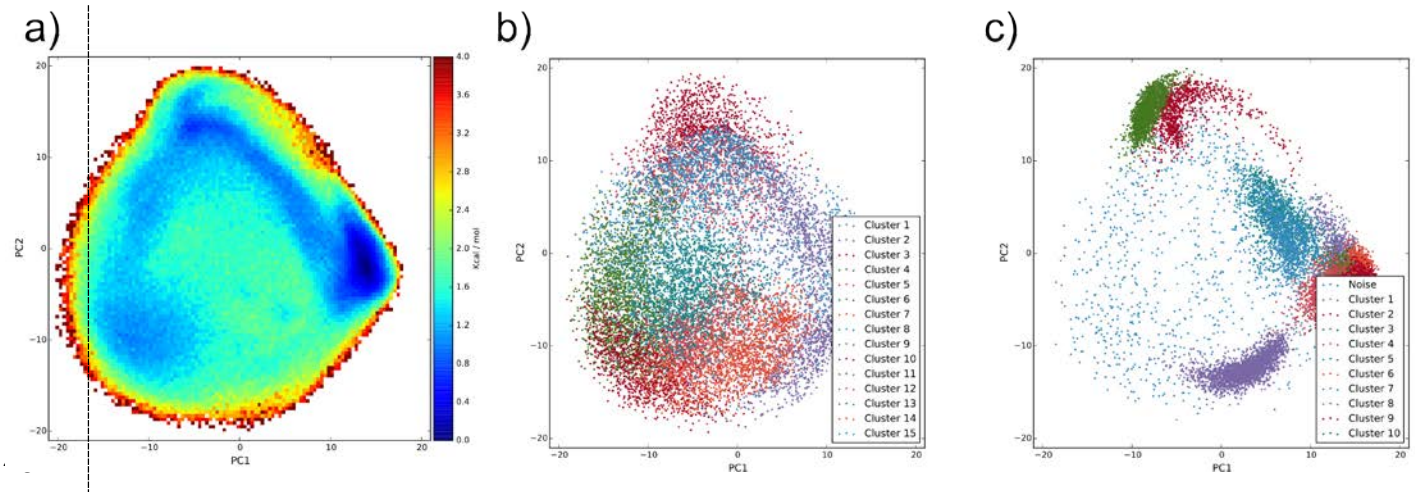


FIGURE 4: PCA projections for the two first (largest) modes of a 1 **μ**s cMD peptide-only simulation. Free energy profiles (a), hierarchical clusters (b) and DBSCAN clusters (c).

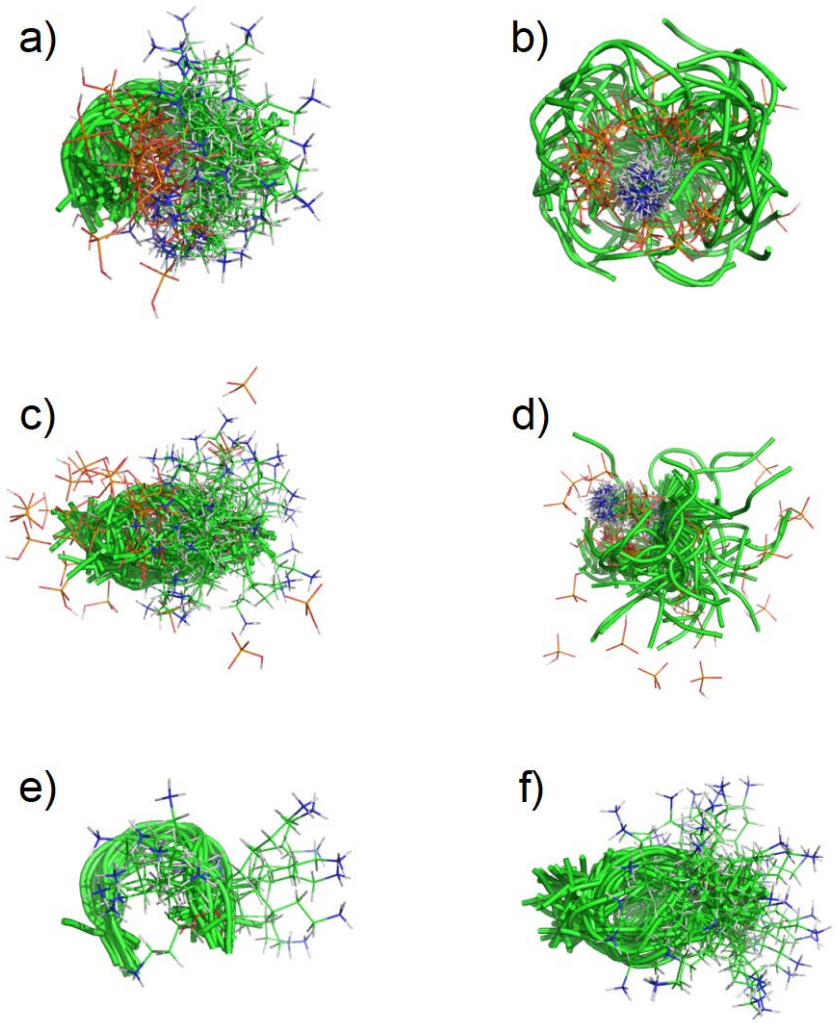
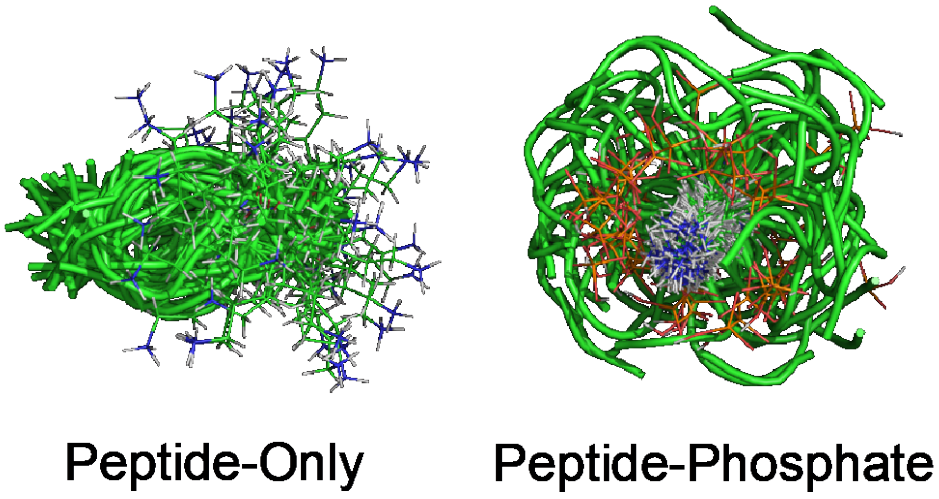


FIGURE 5: Superposition of clusters. DBSCAN clusters (a,b), and hierarchical clusters (c,d) from four 100 ns peptide-HPO_4^{2-} aMD simulations, aligned by the peptide backbone (a,c) and the lysine side chain (b,d). In the bottom row are DBSCAN clusters (e) and hierarchical clusters (f) from two **1** μs peptide-only cMD simulations, aligned by peptide backbone.



473

474

475 TOC ENTRY PICTURE

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the Danish Agency for Science via a grant to the innovation consortium 'Natural Ingredients and New Energy'.

The project also was supported by *IBISS: Industrial Biomimetics for sensing and separation,* a platform funded by the Danish National Advanced Technology Foundation.

Acknowledgment

MFG and CHN wish to acknowledge the support for this work through the Innovation Consortium Natural Ingredients and Green Energy (NIGE)-with sustainable purification technologies financially supported by Danish Agency for Science Technology and Innovation.

Abbreviations

aMD, Accelerated Molecular Dynamics

cMD, Conventional Molecular Dynamics

EBPR, Enhanced Biological Phosphorus Removal

IDP, Intrinsically Disordered Protein

KLD, Kullback-Leibler Divergence

MD, Molecular Dynamics

MM-PBSA, Molecular Mechanics Poisson Boltzmann Surface Area

PCA, Principal Component Analysis

QM, Quantum Mechanics

REMD, Replica Exchange Molecular Dynamics

RMSd, Root Mean Square Deviation

SGLD, Self-Guided Langevin Dynamics

References

(1) Elser, J. J. Phosphorus: A Limiting Nutrient for Humanity? *Curr. Opin. Biotechnol.* **2012**, *23*, 833–838.

(2) Rittmann, B. E.; Mayer, B.; Westerhoff, P.; Edwards, M. Capturing the Lost Phosphorus. *Chemosphere* **2011**, *84*, 846–853.

(3) Cordell, D.; Smit, A. L.; Rosemarin, A. Sustainable Use of Phosphorus. *Plant Res. Int.* **2009**, 1–109.

(4) Kauwenbergh, S. J. Van. World Phosphate Rock Reserves and Resources. *IFDC* http://pdf.usaid.gov/pdf_docs/PNADW835.pdf.

(5) Kimball, S. M. Mineral Commodity Summaries. *USGS* **2015**, http://minerals.usgs.gov/minerals/pubs/mcs/2015/mc.

(6) Gorazda, K.; Wzorek, Z.; Tarko, B.; Nowak, A. K.; Kulczycka, J.; Henclik, A. Phosphorus Cycle — Possibilities for Its Rebuilding. **2013**, *60*, 725–730.

(7) Selman, M.; Greenhalgh, S. Eutrophication: Policies, Actions, and Strategies to Address Nutrient Pollution. *WRI Policy Note* **2009**, http://pdf.wri.org/eutrophication_policies_actions.

(8) Rittmann, B. E.; McCarty, P. L. *Environmental Biotechnology: Principles and Applications*; McGraw-Hill series in water resources and environmental engineering; McGraw-Hill, 2001.

(9) Yuan, Z.; Pratt, S.; Batstone, D. J. Phosphorus Recovery from Wastewater through Microbial Processes. *Curr. Opin. Biotechnol.* **2012**, *23*, 878–883.

(10) Rittmann, B. E. Opportunities for Renewable Bioenergy Using Microorganisms. *Biotechnol. Bioeng.* **2008**, *100*, 203–212.

(11) Poole, K.; Hancock, R. E. W. Phosphate Transport in Pseudomonas Aeruginosa Involvement of a Periplasmic Phosphate-Binding Protein. *Eur. J. Biochem* **1984**, *612*, 607–612.

(12) Wu, H.; Kosaka, H.; Kato, J.; Kuroda, A.; Ikeda, T.; Takiguchi, N.; Ohtake, H. Cloning and Characterization of Pseudomonas Putida Genes Encoding the Phosphate-Specific Transport System. *J. Biosci. Bioeng.* **1999**, *87*, 273–279.

(13) Gruber, M.; Greisen, P.; Junker, C. M.; Hélix-Nielsen, C. Phosphorus Binding Sites in Proteins: Structural Preorganization and Coordination. *J. Phys. Chem. B* **2014**, *118*, 1207–1215.

(14) Walker, J. E.; Saraste, M.; Runswick, M. J.; Gay, N. J. Distantly Related Sequences in the a- and B-Subunits of ATP Synthase, Myosin, Kinases and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. **1982**, *1*, 945–951.

(15) Via, A.; Ferrè, F.; Brannetti, B.; Valencia, A.; Helmer-Citterich, M. Three-Dimensional View of the Surface Motif Associated with the P-Loop Structure: Cis and Trans Cases of Convergent Evolution. *J. Mol. Biol.* **2000**, *303*, 455–465.

(16) Dreusicke, D.; Schulz, G. E. The Glycine-Rich Loop of Adenylate Kinase Forms a Giant Anion Hole. *FEBS Lett.* **1986**, *208*, 301–304.

(17) Pai, E. F.; Krengel, U.; Petsko, G. a; Goody, R. S.; Kabsch, W.; Wittinghofer, a. Refined Crystal Structure of the Triphosphate Conformation of H-Ras p21 at 1.35 A Resolution: Implications for the Mechanism of GTP Hydrolysis. *EMBO J.* **1990**, *9*, 2351–2359.

(18) Saraste M, Sibbald PR, W. A. The P-Loop—a Common Motif in ATP and GTP-Binding Proteins. *Trends Biochem Sci.* **1990**, *11*, 430–434.

(19) Leipe, D. D.; Wolf, Y. I.; Koonin, E. V; Aravind, L. Classification and Evolution of P-Loop GTPases and Related ATPases. *J. Mol. Biol.* **2002**, *317*, 41–72.

(20) Leipe, D. D.; Koonin, E. V.; Aravind, L. Evolution and Classification of P-Loop Kinases and Related Proteins. *J. Mol. Biol.* **2003**, *333*, 781–815.

(21) Watson, J. D.; Milner-White, E. J. A Novel Main-Chain Anion-Binding Site in Proteins: The Nest. A Particular Combination of Phi,psi Values in Successive Residues Gives Rise to Anion-Binding Sites That Occur Commonly and Are Found Often at Functionally Important Regions. *J. Mol. Biol.* **2002**, *315*, 171–182.

(22) Watson, J. D.; Milner-White, E. J. The Conformations of Polypeptide Chains Where the Main-Chain Parts of Successive Residues Are Enantiomeric. Their Occurrence in Cation and Anion-Binding Regions of Proteins. *J. Mol. Biol.* **2002**, *315*, 183–191.

(23) Ramakrishnan, C.; Dani, V. S.; Ramasarma, T. A Conformational Analysis of Walker Motif A [GXXXXGKT (S)] in Nucleotide-Binding and Other Proteins. *Protein Eng.* **2002**, *15*, 783–798.

(24) Theis, K.; Chen, P. J.; Skorvaga, M.; Houten, B. Van; Kisker, C. Crystal Structure of UvrB, a DNA Helicase Adapted for Nucleotide Excision Repair. *EMBO J.* **1999**, *18*, 6899–6907.

(25) Bianchi, A.; Giorgi, C.; Ruzza, P.; Toniolo, C.; Milner-White, E. J. A Synthetic Hexapeptide Designed to Resemble a Proteinaceous P-Loop Nest Is Shown to Bind Inorganic Phosphate. *Proteins* **2012**, *80*, 1418–1424.

(26) Choi, U. B.; McCann, J. J.; Weninger, K. R.; Bowen, M. E. Beyond the Random Coil: Stochastic Conformational Switching in Intrinsically Disordered Proteins. *Structure* **2011**, *19*, 566–576.

(27) Tompa, P. Unstructural Biology Coming of Age. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419–425.

(28) Hilser, V. J.; Thompson, E. B. Intrinsic Disorder as a Mechanism to Optimize Allosteric Coupling in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8311–8315.

(29) Chen, J. Towards the Physical Basis of How Intrinsic Disorder Mediates Protein Function. *Arch. Biochem. Biophys.* **2012**, *524*, 123–131.

(30) D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B., X. W. and P. A. K. *AMBER 14*; San Francisco, 2014.

(31) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712–725.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(33) Bruxelles, U. L. De. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(34) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N· log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(35) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(36) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.

(37) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; Mccammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10*, 2677–2689.

(38) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(39) Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proc. KDD* **1996**, 226–231.

(40) Rajan, A.; Freddolino, P. L.; Schulten, K. Going beyond Clustering in MD Trajectory Analysis : An Application to Villin Headpiece Folding. *PLoS One* **2010**, *5*, 1–12.

(41) Garcia, A. E. Large-Amplitude Nonlinear Motions in Proteins. *Phys. Rev. Lett.* **1992**, *68*, 2696–2700.

(42) Yang, L.-W.; Eyal, E.; Bahar, I.; Kitao, A. Principal Component Analysis of Native Ensembles of Biomolecular Structures (PCA_NEST): Insights into Functional Dynamics. *Bioinformatics* **2009**, *25*, 606–614.

(43) Roe, D. R.; Bergonzo, C.; Cheatham, T. E. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2015**, *118*, 3543–3552.

(44) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. Inference of Structure Ensembles of Flexible Biomolecules from Sparse, Averaged Data. *PLoS One* **2013**, *8*, 1–7.

(45) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492–499.

(46) Kullback, S.; Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.

(47) Kabsch, W. A Discussion of the Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr. Sect. A* **1978**, *34*, 827–828.

(48) Kaufman, L.; Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley Series in Probability and Statistics; Wiley, 1990.

(49) Ankerst, M.; Breunig, M. M.; Kriegel, H. P.; Sander, J. OPTICS: Ordering Points to Identify the Clustering Structure. *Sigmod Rec. (acm Spec. Interes. Gr. Manag. Data), Sigmod Rec* **1999**, *28*, 49–60.

(50) Achtert, E.; Bohm, C.; Kroger, P.; Rothlauf, F. DeLiClu: Boosting Robustness, Completeness, Usability, and Efficiency of Hierarchical Clustering by a Closest Pair Ranking. *Proc. 10th Pacific-Asia Conf. Adv. Knowl. Discov. Data Min.* **2006**, 119–128.

(51) Campello, R. J. G. B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Lect. Notes Comput. Sci* **2013**, *7819*, 160–172.

(52) Campello, R. J. G. B.; Moulavi, D.; Zimek, A.; Sander, J. A Framework for Semi-Supervised and Unsupervised Optimal Extraction of Clusters from Hierarchies. *DATA Min. Knowl. Discov.* **2013**, *27*, 344–371.

(53) Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X. Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. *Data Min. Knowl. Discov.* **1998**, *2*, 169–194.

(54) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(55) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulation. *Protein Chem.* **2003**, *66*, 27–85.

(56) De Visser, S. P.; Quesne, M. G.; Martin, B.; Comba, P.; Ryde, U. Computational Modelling of Oxygenation Processes in Enzymes and Biomimetic Model Complexes. *Chem. Commun. (Camb).* **2014**, *50*, 262–282.

(57) Shaw, B. D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; Mcleavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *SIGARCH Comput. Arch. News* **2007**, *35*, 91–97.

(58) Go, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Grand, S. Le; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(59) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(60) Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 4014–4027.

(61) Baker, C. M.; Best, R. B. Insights into the Binding of Intrinsically Disordered Proteins from Molecular Dynamics Simulation. *WIREs Comput Mol Sci* **2014**, *4*, 182–198.

(62) Tozzini, V. Coarse-Grained Models for Proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.

(63) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381*, 512–518.

(64) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(65) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Algorithms for Molecular Simulations of Biopolymers. *Biopolymers* **2001**, *60*, 96–123.

(66) Nymeyer, M.; S., G.; Garcia, A. E. Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm. *Methods Enzymol.* **2004**, *383*, 119–149.

(67) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109*, 8220–8230.

(68) Cavalli, A.; Ferrara, P.; Caflisch, A. Weak Temperature Dependence of the Free Energy Surface and Folding Pathways of Structured Peptides. *Proteins* **2002**, *47*, 305–314.

(69) Shen, T.; Hamelberg, D. A Statistical Analysis of the Precision of Reweighting-Based Simulations. *J. Chem. Phys.* **2008**, *129*, 034103.

(70) Ceriotti, M.; Brain, G. A. R.; Riordan, O.; Manolopoulos, D. E.; Road, S. P. The Inefficiency of Re-Weighted Sampling and the Curse of System Size in High Order Path Integration. *Proc. R. Soc. A* **2012**, *468*, 2–17.